

# Chapter 26

## Emotional Changes in Japanese Blog Space Resulting from the 3.11 Earthquake

Yukie Sano, Hideki Takayasu, and Misako Takayasu

**Abstract** We quantified the emotional changes observed in social media after major disasters, focusing especially on the Japanese blog space after the Great East Japan Earthquake in 2011. We checked the appearances of Japanese adjectives and found that special emotion adjectives, such as ‘impatient’, and ‘frustrating’ which involve the want to help others but the person has no means to and feels frustrated, occur with considerably increasing frequency. To visualize social mood, we drew a co-occurrence network of adjectives showing a major topological change at the site of the quake. Measuring emotional changes after an emergency has been difficult, but, our research has the potential to achieve it.

### 26.1 Introduction

Social data collected on a large scale has helped to promote our understanding of society and ourselves during this century [1–3]. Drastic changes in information technology enable us to collect various types of communication data, such as those on mobile phone and face-to-face contacts. Hidden patterns of human activities are uncovered, and these results are beginning to be used to solve real social problems.

Web-based data attract many scientists particularly because it is easy to collect data on a large scale for discussion in scientific papers and such data reflect various real social phenomena such as political, market, financial, and daily events.

---

Y. Sano (✉)

University of Tsukuba, Ibaraki, Japan

e-mail: [sano@sk.tsukuba.ac.jp](mailto:sano@sk.tsukuba.ac.jp)

H. Takayasu

Sony Computer Science Laboratories Inc., Tokyo, Japan

Meiji University, Kanagawa, Japan

e-mail: [takayasu@csl.sony.co.jp](mailto:takayasu@csl.sony.co.jp)

M. Takayasu

Tokyo Institute of Technology, Kanagawa, Japan

e-mail: [takayasu@dis.titech.ac.jp](mailto:takayasu@dis.titech.ac.jp)

Examples include Facebook—the world’s largest social networking service [4–6]—Google and Yahoo! search engines [7–9], Wikipedia—a multilingual online encyclopaedia [10–14]—YouTube—an online movie site [15]—and Twitter—a microblogging platform on which users can post messages up to 140 characters [16–22].

Since various large amounts of social media data are available, studies regarding social media have been greatly increasing substantially. In particular, new applications are demanding and attracting, many researches have focused on the connections between real-world phenomena and social media. For example, Mandel et al. focused on the emotional changes in 65,000 tweets during half the month when hurricane Irene approached the U.S. and found that the level of concern had increased depending on region and gender [23]. However, little is known about long-term emotional changes in social media.

There are many ways to detect emotions in texts. Some examples are Linguistic Inquiry and Word Count (LIWC) [24], Positive and Negative Affect Schedule (PANAS) [25], Affective Norms for English Words (ANEW) [26], and Point Of Mood State (POMS) [27, 28]. Furthermore, Emoticons (combination of ‘emotion’ and ‘icon’) such as ‘:)’ and WordNet a thesaurus have been used to detect emotions [29–32]. However, it is known that resulting polarity (i.e., positive or negative) is different depending on the methods [33]. Additionally, most of the methods are English, therefore, there is a not strict dictionary to detect emotions from Japanese texts.

In this paper, we use more than 3.2 billion Japanese blog posts for 6 years since 1 November 2007 as typical Japanese texts. Here we simply use whole Japanese adjectives to detect emotions. Our observation periods include the Great East Japan earthquake in 2011 which is said to have changed social mood qualitatively. First, we describe our data and method in Sect. 26.2. Next we compare the relative frequencies of adjectives before and after the quake and draw co-occurrence networks for the adjectives to visualize social emotions in Sect. 26.3. Summary and discussion are in Sect. 26.4.

## 26.2 Data and Methods

### 26.2.1 Data

We studied Japanese blog data from 1 November 2007 to 31 October 2013 (2192 days) using an Internet service called ‘Kuchikomi@kakaricho’<sup>1</sup> to collect data. This fee-charging service provides an original web-site and an application programming interface (API). The API returns the daily number of blogs in which any given target word appears in a given period. The daily number of blogs containing a blog post

---

<sup>1</sup><http://kakaricho.jp> (Accessed: 11 March 2015).

which includes that the target word occurring more than once is counted. Thus, if one blog post includes a target word multiple times, the API counts it as one. The API also provides a spam filter with three levels—weak, middle and strong—depending on the desired spam detection accuracy. Here, we use the middle level of spam filter to collect data. The full API database contains more than 3.2 billion blog posts from 38 million accounts.

We search for adjectives listed in the MeCab<sup>2</sup>—a Japanese morphological analyzer—dictionary with original surface forms using the API. Because there are various ways of conjugating forms in Japanese, we search for just their original form for simplicity; there are 1741 adjectives in total. Subsequently, we summarize these adjectives' time series in the case that they have the same pronunciations and meanings. Because Japanese uses three different character sets—Hiragana, Katakana and Kanji (Chinese characters)—instead of an alphabet, people tend to use words with different surface forms that have the same pronunciations and meanings. This procedure leads to 839 adjectives. Finally, we removed extremely low- (less than 10 times per day) or high-frequency (more than 100,000 times per day) adjectives, resulting in a total of 550 adjectives' time series.

### 26.2.2 *z-Test for the Quality of Two Proportions*

In this research, we apply a *z*-test for the quality of two proportions to determine whether adjective *i*'s occurrence  $x_i(t)$  differs at different times  $T_0$  and  $T_1$ . The *z*-score of statistic  $z_i$  is calculated as follows:

$$z_i = \frac{\left| \frac{x_i(T_0)}{X(T_0)} - \frac{x_i(T_1)}{X(T_1)} \right|}{\sqrt{X_i(1 - X_i) \left[ \frac{1}{x_i(T_0)} + \frac{1}{x_i(T_1)} \right]}} \quad (26.1)$$

where  $X(t)$  is the total number of blogs at time  $t$  and  $X_i$  is calculated as follows:

$$X_i = \frac{x_i(T_0) + x_i(T_1)}{X(T_0) + X(T_1)} \quad (26.2)$$

The null hypothesis is that  $z_i$  follows a standard normal distribution. Therefore, we can calculate the *p*-value to check the proportions of adjectives in different time periods.

---

<sup>2</sup><https://code.google.com/p/mecab/> (Accessed: 11 March 2015).

### 26.2.3 Partial Correlation Coefficient

To visualize the change in social emotions at the quake, we construct adjective co-occurrence networks for the pre- and post-quake periods. First, we calculate the Pearson's linear correlation coefficient  $r_{ij}$  between the frequency  $y_i(t)$  and  $y_j(t)$  of adjectives  $i$  and  $j$  respectively, with a difference of normalized time series,

$$y_i(t) = \frac{x_i(t+1)}{X(t+1)} - \frac{x_i(t)}{X(t)}. \quad (26.3)$$

Pearson's linear correlation coefficient  $r_{ij}$  between adjectives  $i$  and  $j$  is calculated as follows:

$$r_{ij} = \frac{\sum_t (y_i(t) - \langle y_i \rangle) (y_j(t) - \langle y_j \rangle)}{\sigma_i \sigma_j} \quad (26.4)$$

where  $\sigma_i$  and  $\sigma_j$  are the standard deviations of  $y_i(t)$  and  $y_j(t)$ , respectively. Here, we use three weeks before and after the quake for the comparison, with  $T = 21$  data points. If adjectives  $i$  and  $j$  have a positive correlation at the 0.01 significant level, then the value of  $r_{ij}$  is greater than 0.5487 for  $T = 21$ .

To extract more essential links from the co-occurrence network, we use the partial correlation coefficient  $r_{ij}^k$  as follows:

$$r_{ij}^k = \frac{r_{ij} - r_{ik} \cdot r_{jk}}{\sqrt{1 - r_{ik}^2} \sqrt{1 - r_{jk}^2}}. \quad (26.5)$$

$r_{ij}^k$  is the partial correlation between  $i$  and  $j$  under the fixed condition of  $k$ . Partial correlation is equivalent to the correlation between residuals  $y_i(t)$  and  $y_j(t)$  after the removing correlation between each  $r_{ik}$  and  $r_{jk}$ .

Normally, we can calculate the partial correlation  $D_{ij}$  that is removed by all other adjectives' effects by calculating an inverse correlation matrix. However, we cannot calculate it that way in this case, because we have only  $T = 21$  data points for each time series and there are 550 adjectives (samples). Therefore, we calculate the partial correlation  $D_{ij}$  for each pair of adjectives  $i$  and  $j$  by removing  $k$ 's effect and checking the following condition.

$$D_{ij} = \begin{cases} \max r_{ij}^k & (\text{if } \forall k, r_{ij}^k \geq D_c) \\ 0 & (\text{otherwise}), \end{cases} \quad (26.6)$$

where  $D_c = 0.5613$  at the 0.01 significance level of partial correlation. To stress the significant partial correlation coefficient, we use maximum value of  $r_{ij}^k$  in this research.

## 26.3 Results

### 26.3.1 Adjectives in Emergency and Normal Periods

We focus on relative changes in adjectives before and after the quake by calculating  $z_i$ , as shown in Eq. (26.1). Here, we define the pre-quake period  $T_0$  as three weeks before the quake, from 18 February to 10 March in 2011. Similarly, we define the post-quake period  $T_1$  as three weeks after the quake, from March 12 to April 1 in 2011.

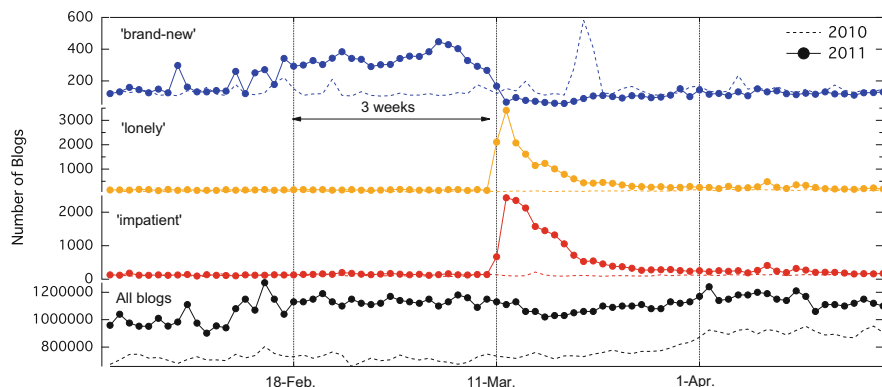
According to the results of the calculation of  $z_i$  for all 550 possible adjectives, 72 adjectives increased and 74 adjectives decreased significantly at the 0.01 significant level. Tables 26.1 and 26.2 show top 10 increased and decreased adjectives respectively. Figure 26.1 shows examples of increased and decreased adjectives.

**Table 26.1** Adjectives that increased significantly ( $p < 0.01$ ) after the quake

#	Word	Original form (Japanese)	$x_i(T_1)/x_i(T_0)$
1	Impatient	はか <sup>カ</sup> ゆい, 歯か <sup>カ</sup> ゆい, 歯痒い	5.91
2	Lonely	こころぼそい, 心細い	5.48
3	Precious	とうとい, 貴い, 尊い	4.84
4	Sorry	こころぐるしい, 心苦しい	4.36
5	Frustrating	もどかしい	4.13
6	Cannot stand the situation	いたたまれない	4.07
7	Heartless	こころない, 心無い	3.57
8	Shameless	あさましい, 浅ましい	3.50
9	Barely afford	ままならない	3.42
10	Base	さかしい	3.40

**Table 26.2** Adjectives that decreased significantly ( $p < 0.01$ ) after the quake

#	Word	Original form (Japanese)	$x_i(T_1)/x_i(T_0)$
1	Brand new	めあたらしい, 目新しい	0.29
2	A little earlier	ひと足早い	0.31
3	Itchy	かゆい, 痒い	0.52
4	Tawdry	やすっぽい, 安っぽい	0.56
5	Cannot wait	まちどおしい, 待ち遠しい, 待遠しい	0.58
6	Salty-sweet	あまからい, 甘辛い	0.59
7	Celebration	めでたい, メデタイ, 目出度い	0.62
8	Bitter	しぶい, 渋い	0.63
9	Spicy	こうばしい, 香ばしい, 芳ばしい	0.63
10	Tough	てごわい, 手ごわい	0.63



**Fig. 26.1** Daily number of blogs including ‘brand-new’, ‘lonely’, and ‘impatient’ from the *top*. Daily number of whole blogs  $X(t)$  is shown in the *bottom*. Solid lines are in 2011 and dashed lines are in 2010 for comparison. Note that sudden increase around 20 March 2010 for ‘brand-new’ is caused by spam blogs, because we confirm that it diminished when we search with the strong level of spam filter

We found that adjectives such as ‘impatient’ which express users’ feelings of frustration have increased considerably according to Table 26.1 (#1, 4, 5, 6). Even these adjectives have different surface forms, albeit similar meanings, e.g. the frustrated feeling of wanting to help others but being unable to do so.

The usage of words such as ‘heartless’ and ‘shameless’ have also increased significantly, according to Table 26.1 (#7, 8, 10). Some people behaved selfishly, buying food and bottles of water under despite the serve shortage conditions after the quake. Therefore, blog posts included complaints about these behaviors with the increased use of these adjectives.

On the other hand, ‘brand-new’, ‘earlier’, and ‘cannot wait’ in Table 26.2 (#1, 2, 5) decreased significantly. These words are often used for positive meanings in expectation of new seasons such as spring and goods such as movies. In fact, many companies canceled or postponed their releases of new product and events. For example, the ceremony of Kyushu Shinkansen opening was canceled, iPad2 (a popular tablet device) release was postponed,<sup>3</sup> and many scheduled movie releases were canceled or postponed.<sup>4</sup> As a result, people lost many chances to use such words.

Furthermore, we found that adjectives related to taste such as ‘salty-sweet’ and ‘bitter’ decreased significantly according to Table 26.2 (#6, 8, 9). These decreases of words may reflect the so-called ‘self-restraint mood’ that people stop to have parties outside such as annual cherry blossom viewing party and to go restaurant. Consequently, the words regarding taste, e.g. restaurant and cook reviews, could

<sup>3</sup>[http://gigazine.net/news/20110316\\_apple\\_delays\\_ipad2\\_launch/](http://gigazine.net/news/20110316_apple_delays_ipad2_launch/) (Accessed: 11 March 2015).

<sup>4</sup><http://www.cinematoday.jp/page/N0031042> (Accessed: 11 March 2015).

decrease. These decreased adjectives seem to be more related to social activities rather than emotions, compared to increased adjectives. Therefore, estimating economic situations by using the adjectives is the crucially interesting future topic.

### 26.3.2 Adjective Changes in the Co-occurrence Network

We constructed co-occurrence network of adjectives during the pre-quake period  $T_0$  and post-quake period  $T_1$ . The node size indicates the relative frequency of words compared with the entire period and the color corresponds to the community it belongs to. The community is decided by the modularity  $Q$  as follows:

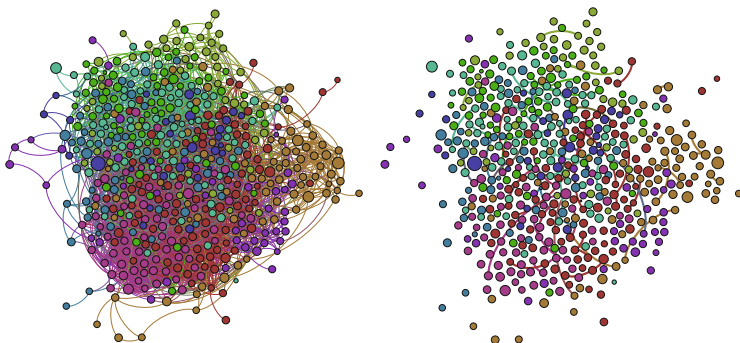
$$Q = \frac{1}{2M} \sum_{i,j=1}^N \left( A_{ij} - \frac{k_i k_j}{2M} \right) \delta(c_i, c_j), \quad (26.7)$$

where  $N = 550$  and  $M$  are total number of nodes and links in network respectively.  $A_{ij}$  is the weight of the link between node  $i$  and  $j$ . Here the weight is correlation coefficient  $r_{ij}$  calculated in Eq. (26.4).  $k_i = \sum_j A_{ij}$  is the sum of the weights linked to node  $i$ , and  $c_i$  is the community which  $i$  belongs to.  $\delta(c_i, c_j)$  is 1 if  $i$  and  $j$  belong to same community ( $c_i = c_j$ ), otherwise 0. In this paper, we maximize  $Q$  for undirected weighted network by a software named Gephi<sup>5</sup> (version 0.8.2) to detect community with the algorithm introduced by Blondel et al.[34].

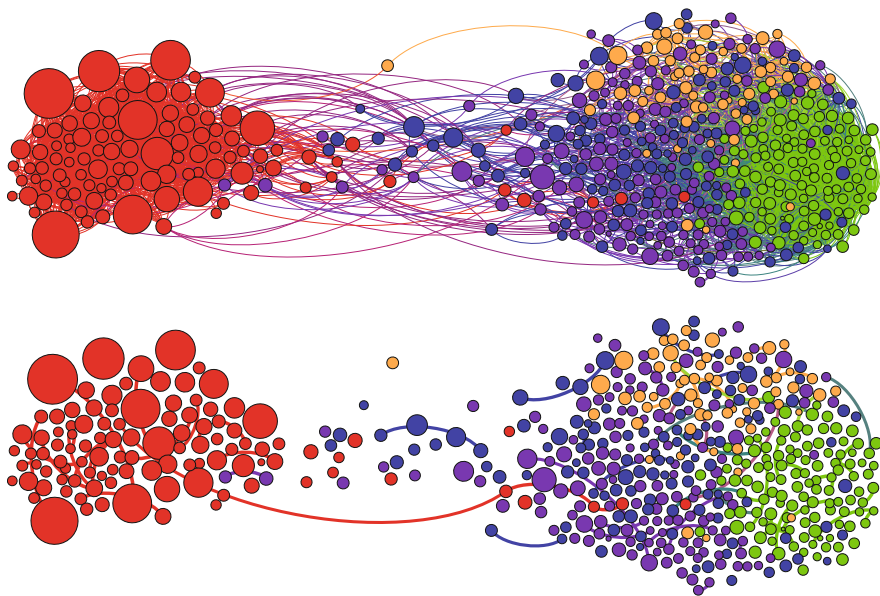
There are 3354 links among 550 nodes during  $T_0$  using the normal correlation coefficient  $r_{ij}$  (Fig. 26.2, left) and 61 links using the partial correlation coefficient  $D_{ij}$  (Fig. 26.2, right). As expected, the node size is nearly the same because there was no major news before the quake  $T_0$  and there are no special properties in these networks. There are 14 communities in the network. The largest community shares 16.36 % and the second shares 14.55 %.

In contrast, there are two major communities post-quake period  $T_1$ . There are 6125 links among the 550 nodes in  $T_1$  by calculating  $r_{ij}$  (Fig. 26.3, top) and 73 links are found by calculating the partial correlation coefficient (Fig. 26.3, bottom). There are five communities in the network. The largest community shares 26.57 % and the second shares 21.59 %. Thus, more nodes categorized into the same communities than  $T_0$  period. This confirms that one community corresponds to the adjectives that increased significantly as shown in the previous section.

<sup>5</sup><http://gephi.org> (Accessed: 11 March 2015).



**Fig. 26.2** Correlation networks consisting of adjectives before the quake  $T_0$ . (*Left*) links are drawn on the basis of the correlation coefficient from Eq. (26.4). (*Right*) links are drawn on the basis of the partial correlation coefficient from Eq. (26.6). Nodes are colored by their community and sized by their relative appearances  $\langle x_i(T_0) \rangle / \langle x_i \rangle$ , where  $\langle x_i \rangle$  is the mean in the entire period and  $\langle x_j(T_0) \rangle$  is the mean in the period  $T_0$



**Fig. 26.3** Correlation networks consisting of adjectives after the quake  $T_1$ . (*Top*) links are drawn on the basis of the correlation coefficient from Eq. (26.4). (*Bottom*) links are drawn on the basis of the partial correlation coefficient from Eq. (26.6). There is one large grouping that consists of increased frequency nodes soon after the quake (*red*). Nodes are colored by their community and sized by their relative appearances  $\langle x_i(T_1) \rangle / \langle x_i \rangle$ , where  $\langle x_i(T_1) \rangle$  is the mean in the  $T_1$  period



## 26.4 Summary and Discussion

We observed social emotions in the Japanese blog space during emergency periods, especially before and after the Great East Japan Earthquake. Here, we focus on the relative changes of adjectives. We found that the feelings such as wanting to help others but having no means and the feeling of frustration increased considerably. Thus, adjectives such as ‘impatient’, ‘sorry’, and ‘frustrating’ increased.

To visualize the change in social mood during the quake, we constructed adjective co-occurrence networks. We found that there is a clear topological difference between the pre- and post-quake periods. We found one large community in the post-quake networks with increased adjectives. This result suggests that people tended to share similar emotions post-quake period.

Our results were derived from a limited case study of Japanese social media during the 3.11 Earthquake. However, our results still have novelty and potential, because no one could record small messages from normal people accurately during emergencies until the advent of social media.

Given that during the quake, rumors and false information are said to have spread via mobile phones and social media [35], investigating social emotions during emergency periods has the potential to provide useful warnings regarding false information diffusion. Because psychologists have observed, on the basis of experiments with limited numbers of subjects, that rumors are more likely to diffuse in an emergency situation [36] and when people feel anxious [37].

We expect that data assimilation—a combination of agent-based simulation and real data analysis—will assist us in preventing potential secondary disasters such as false information diffusion and riots during emergencies. The Internet era has even said to foster ‘digital wildfires in a hyperconnected world [38]’ similar to the spread of real wildfires. Our research sheds light on this universal problem and could issue warning signals of potential digital wildfires. We hope that our results can be applied to prevent such problems in the future.

**Acknowledgements** The authors thank Hottolink, Inc. for providing the data. This study is partially supported by the Grant-in-Aid for Scientific Research (C), Grant Number 24540395 and (B), Grant Number 26310207.

**Open Access** This book is distributed under the terms of the Creative Commons Attribution Non-commercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

1. Watts DJ (2007) *Nature* 445(7127):489
2. Lazer D, Pentland A, Adamic LA, Aral S, Barabási AL, Brewer D, Christakis N, Contractor N, Fowler JH, Gutmann M, Jebara T, King G, Macy MW, Roy D, Alstyn MV (2009) *Science* 323:721
3. Pentland AS (2013) *Sci Am* 309(4):78
4. Onnela JP, Reed-Tsochias F (2010) In: *Proc Natl Acad Sci USA* 107(43):18375
5. Adamic LA, Lento T (2012) In: *Proceedings of the sixth international AAAI conference on weblogs and social media*, pp 371–374
6. Kramer A (2012) In: *Proceedings of the 2012 ACM conference on human factors in computing systems*, pp 767–770
7. Ettredge M, Gerdes J, Karuga G (2005) *Commun ACM* 48(11):87
8. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2009) *Nature* 457(7232):1012
9. Goel S, Hofman J, Lahaie S, Pennock D, Watts DJ (2010) *Proc Natl Acad Sci USA* 107(41):17486
10. Leskovec J, Huttenlocher D, Kleinberg J (2010) In: *Proceedings of the fourth ACM international conference on web search and data mining*, pp 98–105
11. Ratkiewicz J, Fortunato S, Flammini A, Menczer F (2010) *Phys Rev Lett* 104:158701
12. Yasseri T, Sumi R, Kertész J (2012) *PLoS One* 7(1):e30091
13. Moat HS, Curme C, Avakian A, Kenett DY, Stanley HE, Preis T (2013) *Sci Rep* 3:1801
14. Mestyán M, Yasseri T, Kertész J (2013) *PLoS One* 8(8):e71226
15. Crane R, Sornette D (2008) *Proc Natl Acad Sci USA* 105(41):15649
16. González-Bailón S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) *Sci Rep* 1:197
17. Mathiesen J, Angheluta L, Ahlgren PTH, Jensen MH (2013) *Proc Natl Acad Sci USA* 110(43):17259
18. Chew C, Eysenbach G (2010) *PLoS One* 5(11):e14118
19. Sakaki T, Okazaki M, Matsuo Y (2010) In: *Proceedings of the 19th international conference on World Wide Web. ACM, New York*, pp 851–860
20. Sasahara K, Hirata Y, Toyoda M, Kitsuregawa M, Aihara K (2013) *PLoS One* 8(4):e61823
21. Starbird K, Palen L, Hughes A, Vieweg S (2010) In: *Proceedings of the 2010 ACM conference on computer supported cooperative work*, pp 241–250
22. Toriumi F, Sakaki T, Shinoda K, Kazama K, Kurihara S, Noda I (2013) In: *Proceedings of the 22nd international conference on World Wide Web companion*, pp 1025–1028
23. Mandel B, Culotta A, Boulahanis J, Stark D, Lewis B, Rodrigue J (2012) In: *Proceedings of the second workshop on language in social media*, pp 27–36
24. Golder SA, Macy MW (2011) *Science* 333(6051):1878
25. Watson D, Clark LA, Tellegen A (1988) *J Pers Soc Psychol* 54(6):1063
26. Kim E, Gilbert S (2009) Detecting sadness in 140 characters: sentiment analysis and mourning Michael Jackson on Twitter. URL <http://www.webecologyproject.org/>
27. McNair DM, Lorr M, Droppleman LF (1971) *Manual for the profile of mood states*. Educational and Industrial Testing Service, San Diego
28. Bollen J, Mao H, Zeng X (2011) *J Comput Sci* 2:1
29. Lamos V, Lansdall-Welfare T, Araya R, Cristianini N (2013) Analysing mood patterns in the United Kingdom through Twitter content. arXiv:1304.5507v1
30. Hannak A, Anderson E, Barrett LF, Lehmann S, Mislove A, Riedewald M (2012) In: *Proceedings of sixth international AAAI conference on weblogs and social media*, pp 479–482
31. Bertrand KZ, Bialik M, Virdee K, Gros A, Bar-Yam Y (2013) Sentiment in New York city: a high resolution spatial and temporal view. arXiv:1308.5010v1
32. Zhao J, Dong L, Wu J, Xu K (2012) In: *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining*, p 1528

33. Gonçalves P, Benevenuto F, Cha M (2013) PANAS-t: a psychometric scale for measuring sentiments on Twitter. arXiv:1308.1857v1
34. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) J Stat Mech Theory Exp 2008(10):P10008
35. Takayasu M, Sato K, Sano Y, Yamada K, Miura W, Takayasu H (2015) PLoS One 10(4):e0121443
36. Allport GW, Postman L (1948) The psychology of rumor. Henry Holt, New York
37. Walker CJ, Beckerle CA (1987) J Soc Behav Pers 2(3):353
38. World Economic Forum (2013) Global Risks 2013